人工智能算法偏见与健康不公平的成因与对策分析

陈龙，曾凯，李莎，等. 人工智能算法偏见与健康不公平的成因与对策分析［J］. 中国全科医学，2023. ［Epub ahead of print］. DOI：10.12114/j.issn.1007-9572.2023.0007

陈龙[1]，曾凯[2]，李莎[1]，陶璐[1]，梁玮[1]，王皓岑[3]，杨如美[1*]

基金项目：2020 年度国家自然科学基金资助项目（72004098）；2022 年度国家自然科学基金资助项目（72204117）；2020 年度江苏高校哲学社会科学研究一般项目（2020SJA0302）；南京医科大学高层次引进人才项目（NMUR2020006）；南京医科大学研究生优质教育资源建设项目（2021F005）；江苏高校优势学科建设工程项目“护理学”（苏政办发〔2018〕87 号）；“十四五”高等教育科学研究规划课题—网络社交媒体使用中的情绪感染效应对医学生抑郁情绪的影响及干预（苏高教会〔2021〕16 号 YB009）；南京医科大学内涵建设专项护理学优势学科资助。

1.211166 江苏省南京市，南京医科大学护理学院
2.510515 广东省广州市，南方医科大学护理学院
3.47907 美国印第安纳州西拉法叶市，普渡大学护理学院
*通信作者：杨如美，南京医科大学护理人文与管理学系主任，副教授；Email: rumeiyang@njmu.edu.cn

【摘要】 随着信息技术的发展，人工智能为疾病诊疗提供越来越重要的价值。然而，人工智能中存在的算法偏见现象，可导致医疗卫生资源分配不均等问题，严重损害患者的健康公平。算法偏见是人为偏见的技术化体现，其形成与人工智能开发过程密切相关，主要源于数据收集、训练优化和输出应用三个方面。医护工作者作为患者健康的直接参与者，应采取相应措施以预防算法偏见，避免其引发健康公平问题。医护工作者需保障健康数据真实无偏见、优化人工智能的公平性和加强其输出应用的透明度，同时需思考如何基于算法偏见以处理临床实践中的不公平现象，全面保障患者健康公平。该文就健康领域中算法偏见的形成原因和应对策略展开综述，以期提高医护工作者识别和处理算法偏见的意识与能力，为保障信息化时代中的患者健康公平提供参考。

【关键词】 人工智能；算法偏见；健康公平；人为偏见；信息化

The causes and countermeasures of artificial intelligence algorithmic bias and health inequity

CHEN Long[1], ZENG Kai[2], LI Sha[1], TAO Lu[1], LIANG Wei[1], WANG Haocen[3], YANG Rumei[1]
1.School of Nursing, Nanjing Medical University, Nanjing 211166, China
2.School of Nursing, Southern Medical University, Guangzhou 510515, China
3. School of Nursing, Purdue University, Indiana 47906, USA
**Corresponding author: YANG rumei, Director of Nursing Humanity and Management Department of Nanjing Medical University, Associate professor, Email: rumeiyang@njmu.edu.cn*

【Abstract】 With the development of information technology, artificial intelligence shows great potentials for clinical diagnosis and treatment. Nevertheless, algorithmic bias in artificial intelligence can lead to problems such as unequal distribution of healthcare resources, which significantly affect patients’ health equity. Algorithmic bias is a technical manifestation of human bias, which is related to the process of artificial intelligence development, including data collection, model training and optimization as well as output application. Since healthcare providers have a direct impact on patients’ health, they should take measures to prevent algorithmic bias and related health equity. It is also important for healthcare providers to ensure the unbiasedness of health data, optimize the fairness of artificial intelligence, and enhance the transparency of its output application. In addition, healthcare providers also need to consider how to solve algorithmic bias and bias related health inequity in clinical practice in order to fully and properly protect patients' health equity. This paper reviews the causes and countermeasures of algorithmic bias in the health field to improve healthcare providers’ awareness and ability in identifying and addressing algorithmic bias, as well as provide empirical foundations for ensuring the health equity in the information age.

【Key words】 Artificial Intelligence; Algorithmic Bias; Health Equity; Human Bias; Information

近年来，国家提倡加快信息技术与医疗健康行业融合，推动全民健康信息化建设。人工智能（Artificial Intelligence，AI）是健康信息化的重要内容，指计算机利用算法对数据进行学习后，模仿人类的思维和行为[1-2]，从而辅助临床决策。随着技术不断提升，AI 对疾病诊断、监测和治疗的价值日益凸显，在风险因素识别、健康资源分配和精准医疗干预等多方面正发挥重要作用[3]。然而，AI 的学习过程经由人类开发设计，从数据选择、标签确定，到训练优化、审查应用，全过程均涉及人为的选择与决策，即便 AI 本身能够客观反映数据，但其学习到的规则逻辑和社会影响却并非完全客观公正，在辅助临床决策时会产生与人类相似的偏见行为，即算法偏见（Algorithmic Bias），表现为 AI 会因劣势群体的种族、性别、宗教和经济等因素而生成差异的输出结果和健康建议，并对该群体产生不良影响[4]。

信息化时代下，算法偏见普遍存在，且与健康公平（Health Equity）密切相关。健康公平是指个体在改善健康状态时，应拥有公正的机会以获取和利用医疗卫生资源[5]。算法偏见类似人为偏见，会因劣势群体的经济水平或特殊性别特征（如性少数群体）等因素而侵害其健康公平，导致该群体获得较少的医疗资源和诊治机会，健康状态相对较差[1]。Science 杂志指出[6]，基于医疗开销构建的 AI 严重侵害经济劣势群体（如非洲裔种族）的健康公平，从表面上看，医疗开销作为 AI 的训练标签，能够综合反映患者的疾病风险，医疗开销值越高则疾病风险越高，从而准确识别高危风险人群；然而实际上，人为选择的这一标签会导致 AI 算法忽略医疗开销高低背后的本质原因，如经济优势群体容易寻求医疗救治（如门诊就医），产生较多的医疗开销，致使 AI 学习到偏见逻辑，倾向于将经济优势群体判定为高疾病风险人群，最终该群体分配较多医疗资源，影响经济劣势群体的健康公平。

因此，算法偏见的本质并非算法自身问题，而是构建和使用 AI 时所涉及的人为选择和决策问题。由于数据导向的 AI 常被认为比人类思维客观公正，医护工作者在应用 AI 时容易忽略其带来的潜在不公平。鉴于此，本文旨在对健康领域中算法偏见的形成原因和应对策略进行综述，以期提高医护工作者对算法偏见的认识，并提出可能的解决策略，为保障医疗卫生信息化背景下的患者健康公平提供参考。

1 算法偏见的形成原因

算法偏见的形成贯穿于 AI 开发全过程，因数据收集阶段隐含偏见而埋下隐患，因训练优化阶段缺乏公平性而嵌入 AI，因输出应用阶段缺乏透明度而产生影响，并继而加剧数据收集阶段中的偏见，造成算法偏见的恶性循环。明确算法偏见的形成原因有助于医护工作者识别算法偏见现象，预防其对患者健康公平的侵害。

1.1 AI 的健康数据隐含偏见

数据是 AI 学习的核心，也是导致算法偏见的主要原因，理想的健康数据应能够真实无偏见地反映临床实践[7]。然而，人为偏见常常隐藏于临床健康数据之中，尤其是病历记录等非结构化文本数据，致使 AI 学习数据中的偏见，形成算法偏见[8]。例如，Irene 等[9]指出医护工作者常会无意识地认为某类人群不易患有特定疾病，如相比于男性，女性因较少的吸烟行为被认为不易患有肺炎，从而导致在记录女性肺炎患者的症状时，医护工作者倾向于将其症状描述为复杂多样，致使该群体的文本数据异质性较大，难以真实反映疾病情况；若利用该文本数据以预测肺炎患者的死亡率，将形成算法偏见，体现为 AI 对女性肺炎患者的预测结局复杂多样，继而导致医护工作者易对该群体采取错误的医疗决策。

临床情境差异导致的样本缺乏代表性同样会使健康数据隐含偏见。临床罕见疾病或偏远地区人群数据缺失等问题将导致 AI 无法充分学习这一类群体的健康特征，从而引发算法偏见[10]。例如，Marissa 等[11]在通过 AI 预测酒精滥用行为时发现，由于数据中低年龄患者的占比较少，致使 AI 对该人群的疾病诊疗方式学习不充分，易出现错误的预测结果，从而导致该群体中的高风险患者未能及时接受治疗，造成健康公平问题。综上，健康数据常常难以真实反映临床实践，人为无意识偏见和临床样本缺乏代表性等问题可使健康数据隐含偏见，埋下算法偏见的隐患。

1.2 AI 的训练优化缺乏公平性

AI 训练优化的过程由人类开发设计，目前健康领域侧重于使 AI 能够输出与实际情况相一致的诊断或预测结果，然而，准确的输出结果并不意味 AI 能够平等对待劣势群体，即 AI 的精确性并不等同于公平性[12]。具体而言，精确性和公平性均是 AI 训练优化的评估指标，精确性侧重于从样本总体的角度评估 AI 预测结果与实际情况的相符程度，而公平性侧重于从样本亚组的角度评估不同亚组之间 AI 的预测能力（如预测值、假阳性率）有无差异[13]。医护工作者在开发 AI 时若忽略对公平性这一指标的考量，将致使客观公正的 AI 算法不受约束地直接学习数据中的隐含偏见，形成偏见的规则逻辑，导致 AI 区分对待劣势群体[14]。

在训练优化阶段中，医护工作者对 AI 训练方式、训练内容的选取将直接影响公平性。由于不同亚组间（如非洲裔和拉丁裔）患者的基因、生物、社会等因素存在差异，医护工作者若采用基于样本总体的经典训练方式，如决策树模型等，易导致 AI 虽然能够准确地进行健康预测与决策，却会忽略各亚组间的差异，难以同时提升 AI 的精确性和公平性[15]。AI 的输入特征、输出标签等训练内容也可影响公平性。国外多项研究在预测肾小球滤过率时发现，

将种族作为输入特征虽有助于提升 AI 的精确性，却可能由于不同种族之间存在生理差异，致使 AI 的公平性降低，亚组上表现为非洲裔美国人的肾小球滤过率预测值过高，诱导医务工作者将其肾脏功能视为更加健康，从而延误疾病的诊断与治疗[16-17]。综上，AI 的精确性不等同于公平性，医护工作者在选取 AI 的训练方式及训练内容时，若仅关注精确性而忽视公平性，易导致客观公正的 AI 算法学习数据中的偏见，促使算法偏见形成。

1.3 AI 的输出应用缺乏透明度

算法偏见的判定包括两个条件，一是 AI 区分对待劣势群体，二是相较于优势群体，AI 对劣势群体产生实际不良影响，然而，由于 AI 技术具备"仅判相关"和"黑箱"的特性，导致医护工作者在构建和应用 AI 时存在低透明度问题，即难以评估人为选择内容对 AI 逻辑与输出的影响，从而加剧算法偏见[4]。首先，AI 具有"仅判相关"的特性，即 AI 仅能学习数据中变量之间的相关性，而不考虑推理与决策的因果关系，这一特性会导致 AI 从数据中学习超出人类预期的异常逻辑规则[18-19]。例如，Rich 等[20]在预测肺炎患者死亡风险时指出，同时患有肺炎和哮喘的患者多在重症监护室接受治疗，鉴于重症治疗技术和资源的优势分配，使得该群体的死亡率反而低于仅患有肺炎的患者，导致 AI 学习反常的相关性逻辑，错误地将哮喘判定为肺炎患者的保护因素。其次，AI 技术常被视为"黑箱"，即高度复杂的 AI 计算过程通常超出人类的认知与理解能力，致使医护工作者难以明确和解释 AI 输出结果的产生原因[13]；Raja[21]等在此基础上指出，实践应用过程中 AI 的普及推广和临床任务的复杂性等外界因素会再次加剧这一问题，促使医护工作者易默认 AI 的输出结果正确无误，阻碍其发现算法偏见的形成和影响。最终，基于 AI 偏见输出的临床决策将诱发健康公平问题，并在实践应用过程中加剧数据产生与收集的隐含偏见，形成算法偏见的恶性循环[22]。

2 算法偏见的应对策略

面对算法偏见的挑战，医护工作者需保障真实无偏见的数据收集，优化训练过程中的 AI 公平性，并加强 AI 输出应用的透明度，从而减少算法偏见，避免 AI 因患者的种族、性别等差异而对其产生劣势影响，以维护患者的健康公平。

2.1 保障 AI 健康数据的真实无偏见

高质量数据是降低算法偏见的关键，医护工作者作为健康数据的生产者、采集者和管理者，需保障健康数据的真实无偏见[22]。首先，医护工作者需减少健康数据中的人为偏见，可利用标准化数据以规范数据的记录内容，通过对比主客观资料，识别和纠正自身因无意识偏见而对劣势群体做出的差异推断，从而避免录入偏见信息[23]；亦可采用对抗学习等技术手段以选取真实数据，或通过提升标准化客观数据的内容占比，从而减弱健康数据中的人为偏见[10, 24]。其次，医护工作者需了解健康数据的分布情况，分析各亚组数据对 AI 学习的影响，以此判断是否存在劣势群体数据缺失等问题，进而通过数据审查、数据集合并等方法确保不同亚组的数据分布均衡[12, 19, 25]。例如，Ziad[6]等在预测疾病风险人群时，通过调整白种人和非洲裔美国人的预测标签，使两者的健康数据分布相似，最终非洲裔美国人的健康资源分配率从 17.7%提升至 46.5%。此外，医护工作者应知晓健康数据中的偏见本质上来源于临床实践，故需考虑如何从偏见数据中识别和管理人为偏见，以保障真实无偏见的数据产生与收集[10]。

2.2 优化 AI 的公平性

高精确性 AI 虽然能够输出准确的预测结果，却可能因缺乏公平性而区分对待劣势群体，医护工作者需在训练优化阶段中评估和改善 AI 的公平性[26]。AI 公平性的评估可从总体和亚组两个角度切入，医护工作者需结合亚组间患者健康状况差异以选取合适的评估方法[14]。具体而言，当不同亚组间患者健康状况差异较小时，应选择从总体角度评估，判断 AI 对同一总体患者的预测值，如疾病发病率，是否会随患者性别、种族等信息的改动而发生异常变化[13]；反之，当亚组间患者健康状况差异较大时，应从亚组角度评估，判断 AI 在不同亚组间的真阳性率等预测能力有无差异[24]。此外，Pedro 等[15]在亚组角度的基础上指出，评估方法的选取尚需考虑 AI 预测结果对患者健康结局的影响，如当预测结果对患者产生较大的消极影响时，应侧重于评估 AI 的假阳性率，消极影响较小时则侧重于评估假阴性率。

医护工作者可通过调整 AI 的训练方式，纠正 AI 规则逻辑的学习过程，从而改善公平性，降低算法偏见。例如，Marissa 等[11]在预测酒精滥用行为时指出，当 AI 在亚组角度上的预测能力（如假阳性率）存在差异，可将数据按照性别、种族等进行亚组分类，并针对各亚组数据进行逐一训练，一定程度上能够维护 AI 对劣势群体的学习过程，具有改善公平性的潜能；Yan 等[27]进一步指出，当劣势群体的数据严重缺乏，基于亚组的逐一训练效果不佳时，可在训练阶段采用迁移学习的算法，通过利用优势群体的数据，以改善 AI 对劣势群体的学习过程。此外，医护工作者应思考 AI 的训练内容是否客观公正，需侧重于选取能够直接反映个体健康状况的训练内容[10]；例如，对于上述根据医疗开销这一标签构建的 AI，表面上该 AI 能够基于医疗开销的高低，看似公正准确地预测高危风险人群，但

若将该标签替换为慢性共病个数，可发现相同风险下，不同种族之间患者的健康状况存在明显差异[6]。值得注意的是，优化 AI 公平性可能会以降低其精确性为代价，因此医护工作者应权衡如何在提升 AI 公平性的同时维持高精确性，从而保障 AI 的临床应用价值[28]。

2.3 加强 AI 输出应用的透明度

医护工作者作为患者健康的直接参与人员，需能够理解 AI 偏见结果的产生原因，并判断其对患者健康公平的实际影响，从而改善 AI 在输出应用过程中的低透明度问题，及时预防算法偏见[2]。首先，医护工作者需提升对 AI 学习过程的理解，积极参与 AI 开发的全过程，包括数据审查、特征选择、算法构建与优化、性能评估与验证等各开发环节，可辅以因果网络或半监督学习等技术手段协助识别 AI 输入数据和输出结果之间的规则逻辑，使 AI 学习过程有迹可循，从而及时发现 AI 偏见输出的形成[29-31]。其次，医护工作者需深思临床任务的内在特征，明确使用 AI 决策的必要性和现实意义，权衡 AI 决策的优势与局限，以避免过度依赖 AI 的输出结果[2, 21]。再者，医护工作者在应用 AI 过程中，需能够从个体行为、环境交互、社会文化和医疗政策等方面综合分析 AI 输出结果对患者健康公平的影响[32]。这一目标需医护工作者咨询和倾听多方意见以避免片面评估，可联合患者、技术人员和管理人员等利益相关人群共同参与 AI 审查，或通过可视化手段直接呈现 AI 应用对临床效益的影响，从而预防算法偏见的不良影响[13, 23, 28]。然而，提升 AI 输出应用的透明度虽能降低算法偏见的影响，却无法消除人为因素导致的偏见现象，因此医护工作者在分析 AI 对患者健康公平的影响时，需思考如何将算法应用与人为偏见相剥离，实现在审查 AI 的同时能够识别临床实践中的不公平现象。

3 总结与展望

AI 作为一把双刃剑，虽能为临床实践带来重要价值，却会因算法偏见而损害劣势群体的健康公平。算法偏见本质上是人为偏见的技术化体现，而这种偏见常因算法或数据的客观性而被忽略。由于现实生活中存在偏见现象，导致收集的健康数据中隐含偏见，致使 AI 学习到偏见的规则逻辑，继而在临床实践中对劣势群体产生不良影响，形成恶性循环。医护工作者作为 AI 开发设计与实践应用的主要参与人员，应知晓人为选择与决策可能导致客观算法的不公平性，故需具有应对算法偏见的意识与能力。首先，医护工作者需保障健康数据真实无偏见，避免在收集客观数据时加入个人的主观推断，并注意是否存在数据比例失衡、偏远数据缺失等问题，从而确保数据能够准确无误地反映现实世界；其次，医护工作者需权衡 AI 训练优化的公平性与精确性，动态评估人为选择的训练方式和训练内容对 AI 性能所产生的影响，从而确保 AI 在公平的前提下实现最大效益；最后，医护工作者需加强 AI 输出应用的透明度，通过深化自身对临床任务和 AI 技术的认识，评估传统决策与 AI 决策对患者健康的影响，权衡 AI 技术在临床实践中的优劣，从而避免因过度依赖 AI 技术而忽略患者主体，防止 AI 算法偏见产生实际影响，并尚需思考如何辨别和管理临床实践中的不公平现象及人为偏见，全面维护患者健康公平[33]。

作者贡献：陈龙负责论文撰写和修改；曾凯、李莎、陶璐、梁玮、王皓岑参与修改论文；杨如美负责论文选题、修改和质量控制。全部作者均阅读并同意了最终稿件的提交。

本文无利益冲突。

参考文献：

[1] CHEN J, KALLUS N, MAO X, et al. Fairness Under Unawareness: Assessing Disparity When Protected Class Is Unobserved[C]//2019 Proceedings of the conference on fairness, accountability, and transparency (FAT). Atlanta, GA, USA: Association for Computing Machinery, 2019: 339-348.

[2] KARKHAH S, JAVADI-PASHAKI N, FARHADI FAROUJI A, et al. Artificial intelligence: Challenges & opportunities for the nursing profession [EB/OL]. (2022-07-21) [2022-12-19]. https://onlinelibrary.wiley.com/doi/10.1111/jocn.16455.

[3] MHASAWADE V, ZHAO Y, CHUNARA R. Machine learning and algorithmic fairness in public and population health[J]. Nature Machine Intelligence, 2021, 3: 659–666. DOI: 10.1038/S42256-021-00373-4.

[4] 詹好. 大数据时代下数据挖掘中的算法歧视研究[D]. 湖南：湖南师范大学, 2020.

ZHAN H. Research on Algorithm Discrimination of Data Mining in the Age of Big Data[D]. Hunan: Hunan Normal University, 2020.

[5] BRAVEMAN P. Health disparities and health equity: concepts and measurement[J]. Annu Rev Public Health, 2006, 27: 167-194. DOI: 10.1146/annurev.publhealth.27.021405.102103.

[6] OBERMEYER Z, POWERS B, VOGELI C, et al. Dissecting racial bias in an algorithm used to manage the health of populations[J]. Science, 2019, 366(6464): 447-453. DOI: 10.1126/science.aax2342.

[7] KEUROGHLIAN A S. Electronic health records as an equity tool for LGBTQIA+ people[J]. Nature Medicine, 2021, 27(12): 2071-2073. DOI: 10.1038/s41591-021-01592-3.
[8] KARNIK N S, AFSHAR M, CHURPEK M M, et al. Structural Disparities in Data Science: A Prolegomenon for the Future of Machine Learning[J]. The American Journal of Bioethics, 2020, 20(11): 35-37. DOI: 10.1080/15265161.2020.1820102.
[9] CHEN I Y, SZOLOVITS P, GHASSEMI M. Can AI Help Reduce Disparities in General Medical and Mental Health Care?[J]. AMA J Ethics, 2019, 21(2): 167-179. DOI: 10.1001/amajethics.2019.167.
[10] PARIKH R B, TEEPLE S, NAVATHE A S. Addressing Bias in Artificial Intelligence in Health Care[J]. JAMA, 2019, 322(24): 2377-2378. DOI: 10.1001/jama.2019.18058.
[11] BORGESE M, JOYCE C, ANDERSON E E, et al. Bias Assessment and Correction in Machine Learning Algorithms: A Use-Case in a Natural Language Processing Algorithm to Identify Hospitalized Patients with Unhealthy Alcohol Use[J]. AMIA Annu Symp Proc, 2022, 2021: 247-254.
[12] CHEN I Y, PIERSON E, ROSE S, et al. Ethical Machine Learning in Healthcare[J]. Annu Rev Biomed Data Sci, 2021, 4: 123-144. DOI: 10.1146/annurev-biodatasci-092820-114757.
[13] LEPRI B, OLIVER N, LETOUZÉ E, et al. Fair, Transparent, and Accountable Algorithmic Decision-making Processes[J]. Philosophy & Technology, 2018, 31: 611-627. DOI: 10.1007/s13347-017-0279-x.
[14] SALEIRO P, KUESTER B, STEVENS A, et al. Aequitas: A Bias and Fairness Audit Toolkit[Z/OL]. (2019-04-29) [2022-12-19]. https://arxiv.org/abs/1811.05577.
[15] KLEINBERG J M, MULLAINATHAN S, RAGHAVAN M. Inherent Trade-Offs in the Fair Determination of Risk Scores[C]//2017 8th Conference on Innovations in Theoretical Computer Science (ITCS). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany: Leibniz International Proceedings in Informatics (LIPIcs), 2017: 1-23.
[16] ANDERSON A H, YANG W, HSU C Y, et al. Estimating GFR among participants in the Chronic Renal Insufficiency Cohort (CRIC) Study[J]. Am J Kidney Dis, 2012, 60(2): 250-261. DOI: 10.1053/j.ajkd.2012.04.012.
[17] LEVEY A S, HOCINE T, TITAN S M, et al. Estimation of Glomerular Filtration Rate With vs Without Including Patient Race[J]. JAMA Internal Medicine, 2020, 180(5): 793-795. DOI: 10.1001/jamainternmed.2020.0045.
[18] MEHRABI N, MORSTATTER F, SAXENA N, et al. A Survey on Bias and Fairness in Machine Learning[J]. ACM Computing Surveys, 2021, 54(6): 1-35. DOI: 10.1145/3457607.
[19] GIANFRANCESCO M A, SUZANNE T, JINOOS Y, et al. Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data[J]. JAMA Internal Medicine, 2018, 178(11): 1544-1547. DOI: 10.1001/jamainternmed.2018.3763.
[20] CARUANA R, LOU Y, GEHRKE J, et al. Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission[C]//2015 Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD). Sydney, NSW, Australia: ACM, 2015: 1721-1730.
[21] PARASURAMAN R, MANZEY D H. Complacency and bias in human use of automation: an attentional integration[J]. Hum Factors, 2010, 52(3): 381-410. DOI: 10.1177/0018720810376055.
[22] CHU C H, NYRUP R, LESLIE K, et al. Digital Ageism: Challenges and Opportunities in Artificial Intelligence for Older Adults[J]. Gerontologist, 2022, 62(7): 947-955. DOI: 10.1093/geront/gnab167.
[23] MOSS K O, HAPP M B, BRODY A. Nurses' Role in Reducing Inequities for the Seriously Ill[J]. J Gerontol Nurs, 2022, 48(8): 3-5. DOI: 10.3928/00989134-20220629-01.
[24] KAIROUZ P, LIAO J, HUANG C, et al. Generating Fair Universal Representations using Adversarial Models[J]. IEEE Transactions on Information Forensics and Security, 2022, 17: 1970-1985. DOI: 10.48550/arXiv.1910.00411.
[25] KOH P W, LIANG P. Understanding Black-box Predictions via Influence Functions[C]//2017 Proceedings of the 34th International Conference on Machine Learning (ICML). Sydney, NSW, Australia: PLMR, 2017: 1885-1894.
[26] The Lancet Digital Health. New resolutions for equity[J]. Lancet Digit Health, 2022, 4(1): 1. DOI: 10.1016/S2589-7500(21)00280-6.
[27] GAO Y, CUI Y. Deep transfer learning for reducing health care disparities arising from biomedical data inequality[J]. Nat Commun, 2020, 11(1): 5131. DOI: 10.1038/s41467-020-18918-3.
[28] ZHANG Y, BELLAMY R, VARSHNEY K R. Joint Optimization of AI Fairness and Utility: A Human-Centered Approach[C]//2020 AAAI/ACM Conference on AI, Ethics, and Society (AIES). New York, NY, USA: Association for Computing Machinery, 2020: 400-406.
[29] ZHOU Y, LI Z, LI Y. Interdisciplinary collaboration between nursing and engineering in health care: A scoping review[J]. Int J Nurs Stud, 2021, 117: 103900. DOI: 10.1016/j.ijnurstu.2021.103900.
[30] ZHANG L, WU Y, WU X. A causal framework for discovering and removing direct and indirect discrimination[C]//2017 26th International

Joint Conference on Artificial Intelligence. Melbourne, Australia: AAAI Press, 2017: 3929-3935.

[31] PRYZANT R, YANG Z, XU Y, et al. Automatic Rule Induction for Efficient Semi-Supervised Learning[Z/OL]. (2022-10-14) [2022-12-19]. https://arxiv.org/abs/2205.09067.

[32] RICHARDSON S, LAWRENCE K, SCHOENTHALER A M, et al. A framework for digital health equity[J]. NPJ Digital Medicine, 2022, 5: 1-6. DOI: 10.1038/s41746-022-00663-0.

[33] Nature Machine Intelligence. Striving for health equity with machine learning[J]. Nature Machine Intelligence, 2021, 3: 653. DOI: 10.1038/s42256-021-00385-0.